

Entrevista a Javier Gutiérrez Puebla

Mariana Relli

E: entrevistadora Mg. Mariana Relli

JGP: Javier Gutiérrez Puebla

E: Desde Geograficando nos interesó conversar con usted para que nos hable un poco sobre el Grupo tGIS, cómo se compone, su trayectoria, cuáles son sus preocupaciones centrales, las líneas de trabajo que tienen en marcha y sus resultados.

JGP: En la Universidad Complutense hay grupos de investigación que se constituyen de forma más o menos espontánea, son profesores e investigadores que trabajan unos temas, deciden trabajar juntos y esa relación se formaliza en forma de grupos de investigación. Nosotros llevábamos mucho tiempo trabajando en temas de sistemas de información geográfica y transporte, ese es nuestro punto fuerte, desde principios de los 90 empezamos a trabajar con SIG, antes trabajábamos temas de transporte y vimos que esta herramienta era muy útil para seguir trabajando en el mismo campo. Ahí surgió el grupo, al principio éramos pocos, ahora somos más y, con cargo a proyectos, vamos consiguiendo investigadores, tenemos algunos investigadores posdoctorales, otros predoctorales y el foco principal básicamente sigue siendo el mismo: SIG, transporte, movilidad y la vida en la ciudad, el pulso de la ciudad, mejor dicho.

Como grupo tenemos un vínculo especial con Transit, que es el Centro de Estudios de Transporte de la Universidad Politécnica de Madrid. Nuestro grupo de transporte y Transit nos integramos y tenemos un centro común, ahora somos UCM - UPM, eso lo hicimos en el contexto del Campus de Excelencia de Moncloa, un proyecto del Ministerio de Ciencia y Tecnología de hace ya unos años, que lo que buscaba era generar sinergias, hicimos la integración y colaboramos de forma frecuente con este centro de investigación.

Últimamente nos hemos interesado mucho por el Big Data porque hemos visto que ahí hay unas fuentes de datos que nos permiten hacer cosas nuevas, ver cosas que antes no veíamos y, por lo tanto, medir, analizar, modelizar y predecir fenómenos con ayuda de estas nuevas fuentes de datos. Utilizamos nuevas tecnologías para procesar estas fuentes de datos porque tienen varias características; una de ellas, como su nombre indica, es el tamaño, son ficheros masivos, datos masivos, muy grandes, pero luego también porque son datos variados, no son datos que se puedan procesar con gestión de bases de datos tradicional sino que hay que utilizar otros gestores de datos y a veces también, para ganar tiempo, computación paralela.

E: Dentro del campo del transporte, ¿cuáles son sus preocupaciones, en qué líneas están trabajando?

JGP: Nuestras preocupaciones fundamentales han sido la accesibilidad y la movilidad. Accesibilidad entendida como la facilidad para acceder a los lugares deseados por parte de la población o por parte de las empresas. Lo llevábamos analizando desde hace mucho tiempo con SIG y lo hacíamos de forma estática, lo utilizábamos por ejemplo para estudiar el impacto que puede tener una determinada actuación de transporte o el impacto que puede tener un plan de infraestructuras. Con un plan de infraestructuras, la nueva infraestructura produce un ahorro de tiempo y un ahorro de costos de transporte, y ese ahorro se difunde por la red de transporte de forma que los efectos de un determinado tramo que es mejorado o una línea que es mejorada, afectan a espacios que son relativamente distantes, y la única forma de modelizar esto es utilizando redes y sistemas de información geográfica. Hemos hecho para la administración numerosos estudios de este tipo, para la Comunidad Autónoma de Madrid y otras comunidades autónomas, también

para el gobierno central hemos hecho análisis de líneas de alta velocidad o planes de transporte en su conjunto, y también hemos analizado planes de la red transeuropea de carreteras o de ferrocarriles, también a escala europea en su conjunto.

Lo que hemos visto ahora es que con Big Data podemos hacer los análisis en forma dinámica, mientras que antes teníamos una visión estática, ahora podemos ver que los cambios o simplemente las circunstancias de la accesibilidad se modifican durante el día, por ejemplo, en función de las condiciones de tráfico, si hay congestión en la mañana o no la hay en hora valle, en el caso del transporte público, en función de las frecuencias y los horarios, hay frecuencias más altas por la mañana temprano, en hora punta, y frecuencias más bajas en hora valle, pero no solamente en función de cómo cambian las condiciones del transporte, sino también en función de los destinos deseados, que también cambian con el tiempo: por la mañana temprano los destinos deseados son básicamente los lugares de trabajo, mientras que a última hora de la tarde, los destinos deseados son otros y fundamentalmente lo que quiere la gente es ir a hacer compras o a actividades de ocio, de forma que con Big Data podemos modelizar las dos cosas y utilizar modelos de accesibilidad doblemente dinámicos.

La otra línea es la de movilidad. Con nuevas fuentes de datos como los de Big Data podemos analizar la movilidad de forma dinámica y verla en tiempo real, no como antes que teníamos fotos fijas: cómo se mueve la gente de acuerdo a una encuesta domiciliaria que ha hecho la administración en un año determinado. Con estas nuevas fuentes tenemos datos siempre actualizados e incluso podemos monitorizar cambios que se producen en la movilidad. Antes utilizábamos encuestas domiciliarias origen-destino o encuestas domiciliarias de movilidad, donde se elige una muestra de la población y en sus casas se les pregunta sobre qué hicieron el día anterior. Esta es una operación cara, que se hace normalmente en las ciudades grandes cada siete u ocho años, en las ciudades pequeñas no se hace; ahora tenemos la ventaja de que podemos tener datos actuales para ciudades grandes, continuas en el tiempo, pero además también para ciudades medias y pequeñas, donde era más difícil recopilar este tipo de datos.

Y las nuevas fuentes de datos pueden ser muy variadas, desde telefonía móvil, datos de Twitter, dónde está twitteando la gente en cada momento del día, y en el caso de la movilidad de turistas, esta movilidad no es captada por las encuestas tradicionales porque son encuestas a residentes, mientras que en cambio las nuevas fuentes de datos, como por ejemplo las redes sociales tipo Flickr, que contiene fotografías geolocalizadas, o del tipo de Foursquare, que contiene patrones asociados al consumo, o incluso del tipo de Twitter, que está relacionado en el caso de los turistas con los lugares donde se alojan, al final de la jornada llegan al hotel y empiezan a twittear, entonces, con estas nuevas fuentes de datos tenemos información rica sobre los turistas. Ahora estamos en vías de conseguir cerrar un acuerdo, que lo tenemos ya muy avanzado, para utilizar datos de tarjetas bancarias, que con ellas tenemos el consumo de la población residente y también de los turistas, lo cual nos da información realmente útil.

E:¿Podrías explicarnos qué es esto llamado Big Data y de qué manera es aplicado a los estudios geográficos?

JGP: Se suele hablar del Big Data en relación a su carácter masivo, son datos masivos porque son generados de dos formas, por una parte, sensores, cámaras que registran la actividad humana, por ejemplo, la gente que pasa por delante de determinada tienda, hay sensores que están midiendo cuánta gente pasa por delante, o cámaras fijas en un lugar de la ciudad que están registrando cuánta gente entra en determinado recinto, esos son datos de sensores. Pero también hay otros dispositivos como los teléfonos móviles, se puede también captar a los teléfonos móviles a través de las redes wifi, se puede captar la actividad humana a través de tarjetas inteligentes de transporte, el usuario va al metro o va al autobús y queda registrado quién lo va a utilizar, pero también queda registrado cuando utiliza una bicicleta de uso compartido, de estas que tenemos en las ciudades, y en general también en todos los sistemas que se conocen como sistemas de servicios a la movilidad, como por ejemplo los servicios que están surgiendo del tipo de Car To Go, que son compañías de medios de transporte compartido, tú llegas a determinado lugar, utilizas un coche y lo dejas en otro, utilizas una moto o una bicicleta y la dejas en otro lugar, con todo esto dejamos nuestra huella digital, pero también dejamos

nuestra huella digital por otros muchos tipos de sensores y dispositivos. Y luego también cuando utilizamos internet estamos dejando la huella digital, cuando hacemos búsquedas en Google o cuando utilizamos redes sociales como Facebook o Twitter, todo eso crea una enorme cantidad de datos que va creciendo en tiempo real a una velocidad enorme y es lo que conocemos como Big Data.

Para nosotros que trabajamos en Geografía Humana, la gran ventaja es que esa huella digital nos permite analizar la actividad humana, no nos interesa la huella digital en sí misma, sino por aquello que nos pueda indicar con respecto a multitud de cuestiones como por ejemplo la presencia de la población en distintas partes de la ciudad, con el consumo, con la movilidad, el turismo, y con gran cantidad de actividades o de cuestiones relacionadas con la Geografía Humana.

Para caracterizar al Big Data se suele aludir a las tres *V* o las cinco *V*¹: una característica es el volumen, el carácter masivo de los datos; otra es la variedad, estábamos acostumbrados antes a trabajar con datos que puedes mostrar en forma de tabla, pero aquí tenemos datos que pueden ser textos (un *twitt* es un texto, un mensaje es un texto) y hay que analizarlos de una manera distinta, con software de inteligencia semántica que nos permita hacer un análisis de esa información cualitativa de forma rápida. Pero también podemos aludir a una tercera *V* que es la velocidad con la que se generan estos datos, que nos da ese carácter de información continua y que podemos monitorizar en tiempo real, lo cual también está relacionado con el carácter masivo, si los generas a mucha velocidad, estás generando muchos datos. Esas son las 3 principales características del Big Data, las tres *V*, aunque algunos autores hacen alusión a las cinco *V*, a las siete *V*, añadiendo otras características más, pero estas serían las más determinantes.

E: ¿Y qué limitaciones o riesgos encuentran en el uso del Big Data como fuente de datos para los estudios geográficos? ¿Están viendo cosas con las que hay que tener cuidados especiales?

JGP: En la pregunta aparecen dos cuestiones, la primera tiene que ver con la privacidad. La privacidad puede verse violentada con el uso de datos que nosotros estamos aportando, pero los aportamos sin tener conocimiento de lo que después se va a hacer con esos datos. Hay que tener en cuenta que muchos de los servicios que utilizamos son servicios gratuitos y no nos fijamos que estamos pagando esos servicios con los datos. Y tampoco nos fijamos que, cuando hacemos la descarga de una *app*, muchas veces hay una serie de condiciones y nosotros le decimos que sí a todo; ahora en Europa hay un nuevo reglamento de protección de datos y en estas semanas estamos recibiendo multitud de mensajes diciendo si queremos aprobar o no las condiciones, ese es un paso adelante. Yo creo que la privacidad es importante, pero quizás, al menos aquí en Europa, hemos pasado de una situación en la que decíamos que sí a todo con gran facilidad y no nos preocupaba la privacidad a otra que ahora estamos enormemente preocupados por eso. Creo que hay que buscar equilibrios, porque si todo el mundo quiere que se mantenga la privacidad y no autoriza que los datos sean aprovechados para estudios, las ventajas del Big Data para hacer estudios en investigación, pero también por parte de empresas, que puedan ser estudios lícitos, se podrían perder, y también para las Smart Cities, las Ciudades Inteligentes se alimentan en buena parte de estos datos masivos.

Tenemos muchas ventajas porque el Big Data nos ofrece muchas potencialidades, pero es cierto también que el Big Data tiene limitaciones importantes. Siempre que se habla de Big Data, la primera limitación que aparece es el sesgo, que es una información que está sesgada y que nosotros difícilmente podemos controlar. Cuando hacemos una encuesta, extraemos una muestra de forma que sea representativa de toda la población; cuando trabajamos con Big Data, lo que tenemos es lo que tenemos, sí que es cierto que nosotros después podemos tomar alguna medida para eliminar una parte de la información, una parte de los datos que sepamos que no nos aportan nada, que son basura, hay que hacer limpieza de datos, como tenemos muchos datos, podemos limpiar mucho, que tendremos todavía mucho, igual tendremos una masa crítica muy importante, hablamos de miles de millones de registros, pero aun así, esa parte de los datos que nos quedamos puede tener un cierto nivel de sesgo.

Pero no se puede calificar al Big Data en el sentido de que es información sesgada de forma generalizada porque el nivel de sesgo varía mucho, por ejemplo, hay redes sociales que sabemos que son utilizadas por muy

poca gente, entonces sabemos que ahí el sesgo es mayor, pero al teléfono móvil lo utiliza prácticamente toda la población en la mayor parte de las ciudades del mundo, sobre todo en las de los países más desarrollados, y nos podemos encontrar con compañías que tienen una cuota de mercado del 30 ó 40 %, si tenemos el 30 % y no hay razones para pensar que esa compañía está representada de forma no igualitaria en los distintos segmentos de la sociedad, tenemos una muestra enorme, francamente buena, y con un nivel de sesgo mínimo. También las encuestas tienen al final sesgo, porque todos sabemos que hay personas que queremos encuestar y nos dicen que no repetidamente, y eso introduce ya un sesgo, y aunque hayamos hecho un diseño de muestra estratificada, el porcentaje de viejos, de niños, etc., pero al final, los que no responden nos introducen un sesgo. El sesgo es siempre inherente a los trabajos en ciencias sociales cuando no trabajamos con toda la población y eso es casi siempre, solamente trabajamos con toda la población con censos y poco más, e incluso en los censos tampoco está toda la población, pero ese nivel de sesgo es muy variable.

E: Mencionas en un artículo que hay una tendencia a reemplazar preguntas del censo por el uso de esta tecnología, ¿es algo que ya está sucediendo o que se acerca?

JGP: Los censos son operaciones carísimas para los países y, por otra parte, uno tiene la idea de que el censo cubre el 100 % de la población, y también que el censo de viviendas cubre el 100 % de las mismas y que es una información enormemente fiable. Sí que es muy fiable, pero al final sabemos que no cubre a toda la población, no cubre a todas las viviendas, y tiene errores. Es muy caro y contiene errores, entonces, en función del precio y de que en los países desarrollados nos encontramos con que la administración tiene una buena parte de la información que se está preguntando en los censos (y aquí, en Europa y en España, tenemos por norma que no debemos preguntar por información que ya tenemos en la administración, si usted ya la tiene, para qué vuelve a preguntar), los países europeos se han puesto de acuerdo para que en el próximo censo, que es el del año 2021, no se haga una operación censal tradicional, sino a partir de los registros administrativos configurar una información que sea similar a la del censo. Registros administrativos hay por todas partes, en distintas oficinas, por ejemplo, en las oficinas de la seguridad social, tienen muchísima información, lo mismo ocurre con los datos de los ayuntamientos, de los padrones de población, los padrones de vehículos, hay multitud de información que tiene la administración y que, si se junta, conforma algo que es muy parecido a un censo en cuanto a la cantidad de información que puede suministrar. Es cierto que alguna información no sale de allí y es en esos casos en los que se está pensando utilizar el Big Data, por ejemplo, en el caso del censo de viviendas de España se divide a las viviendas según vivienda habitual, vivienda secundaria y vivienda desocupada, y en los registros de viviendas no podemos saber si estamos en cualquiera de esas tres situaciones, pero es muy fácil con los datos de consumo de energía eléctrica que tienen los hogares, saber si estamos en una u otra categoría, saber si están consumiendo energía eléctrica todo el año, todos los días del año, está claro que son viviendas ocupadas, si el consumo es puntual, fines de semana o vacaciones, está claro que es segunda residencia, pero si no hay consumo, está claro que es una vivienda desocupada. Lo mismo ocurre con datos de movilidad, en los últimos censos españoles se ha ofrecido información muy utilizada por las universidades, por los investigadores, se ha ofrecido información sobre movilidad de la población: lugar de trabajo, lugar de estudios, el medio de transporte utilizado y el tiempo de viaje; a eso no lo podemos sacar por registros administrativos, entonces, la idea de varios organismos que hacen los censos en los países europeos es utilizar datos de compañías de telefonía móvil para hacer esas matrices de viajes y hacer esos cálculos.

Es un acuerdo europeo y no es algo nuevo del todo, ya en el censo de 2011 no se hizo una operación censal completa en la mayor parte de los países europeos, inclusive en España, y en el caso de España se hizo una muestra del 10 % de la población, pero ya no se hizo una operación censal. Para el 2021 ya ni siquiera se va a hacer una muestra porque se va a cambiar la sistemática, en coordinación con otros países europeos, cada país es libre de hacer lo que quiera, pero se trata de coordinar las acciones para hacer cosas parecidas.

E: ¿Se han establecido vínculos entre las líneas de investigación del Grupo tGIS y la formación en la Licenciatura en Geografía en la UCM?

JGP: Nosotros fuimos pioneros en la implantación de SIG en planes de estudios de licenciatura, ya en el plan de 1993 implementamos un conjunto de asignaturas de SIG y de teledetección y, además, desde 1992 teníamos un posgrado de SIG y teledetección. Después dejamos de impartir ese posgrado y lo que hicimos fue impartir un Máster en Tecnologías de la Información Geográfica, oficial, con título propio; es decir, tenemos enseñanza de SIG, que es la herramienta principal de nuestro grupo, no solamente en el grado sino en nuestro máster y en otros, por ejemplo, un Máster en Estadísticas Oficiales que impartimos varios departamentos de la Universidad Complutense y ahí tenemos una asignatura de SIG y estadística espacial, también en el Máster de Arqueología y en otros. La herramienta de SIG ha tenido mucha demanda en estudios de grado y posgrado y ahora hay oferta de asignaturas en varios másteres y en nuestro grado.

En cuanto a Big Data, esto es algo mucho más nuevo y hay que hacer reformas de los planes de estudios para poder introducirlo en el grado, pero sí en el máster hemos ido introduciendo ya de alguna forma algo de aprendizaje con Big Data, enseñando a los estudiantes qué es y en algunos trabajos de fin de máster ellos ya están trabajando con data-sets obtenidos a través de tecnologías Big Data.

E: ¿Y esta introducción significa dejar de lado otros contenidos que no se consideran necesarios?

JGP: Desde el principio de este posgrado somos conscientes de que un año es muy poco, debería ser de dos años, y a pesar de todo estamos incluyendo constantemente nuevos contenidos, lo cual nos complica muchísimo más poder comprimir esas novedades. No solamente tenemos estas novedades del Big Data, que nos obliga a hacer reestructuración, por ejemplo, hemos eliminado una asignatura de cartografía que ya no era muy importante y no era muy demandada por los estudiantes y eso nos ha permitido expandir un poco otras asignaturas con mucha demanda, y nos dio juego para meter nuevos aspectos relacionados no solamente con cambios en los datos, como el Big Data, sino otros cambios en tecnologías como por ejemplo los drones, que antes no teníamos en el plan de estudios porque antes no se utilizaban los drones de la manera que se utilizan ahora. Con los drones estamos consiguiendo imágenes de una altísima resolución espacial y en tiempo real, y además el proceso de esos datos para conseguir imágenes en tres dimensiones es rapidísimo y todo ese proceso lo estamos implementando en el máster. Y también otros datos, como por ejemplo los de LIDAR que son imágenes de radar, que antes eran mucho menos frecuentes, pero ahora el Instituto Geográfico Nacional ofrece estos datos para ser descargados desde su página web, desde la infraestructura de datos espaciales del Instituto Geográfico Nacional, y nos da una resolución muy alta para las coordenadas X, Y, Z, y nos permite hacer una clasificación de usos del suelo, lo cual, para trabajar con la ciudad esas imágenes son interesantísimas, pues, viendo que teníamos estas imágenes, cuando tuvimos la oportunidad hicimos un cambio y las introdujimos. En este máster estamos constantemente introduciendo cambios.

E: Muchas gracias Javier

MÁS INFORMACIÓN:

tGIS: Grupo de Investigación Transporte, Infraestructura y Territorio, Universidad Complutense de Madrid: <https://www.ucm.es/tgis/presentacion>

Transit: Centro de Investigación del Transporte, Universidad Politécnica de Madrid: <http://www.transyt.upm.es/index.php/es>

Máster en Tecnologías de la Información Geográfica, UCM: <https://www.ucm.es/estudios/master-informaciongeografica>

PARA SEGUIR LEYENDO:

Gutiérrez Puebla, J. (2018). Big Data y nuevas geografías: la huella digital de las actividades humanas. *Documents d'Anàlisi Geogràfica*, vol. 64/2. Disponible en: <http://dag.revista.uab.es/article/view/v64-n2-gutierrez>

NOTAS

1 Las 5 V refieren a: volumen, velocidad, variedad, veracidad y valor (Gutiérrez Puebla, 2018).